

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NO. PHN 17,717
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. Application No. (if known, see 37 CFR 1.5) <b>09/869386</b>
INTERNATIONAL APPLICATION NO. PCT/EP00/10713	INTERNATIONAL FILING DATE 10/27/00	PRIORITY DATE CLAIMED 10/29/99
TITLE OF INVENTION ELIMINATION OF NOISE FROM A SPEECH SIGNAL		
APPLICANT(S) FOR DO/EO/US CHAO-SHIH J. HUANG		
<p>Applicant(s) herewith submit to the United States Designated/Elected Office (DO/EO/US) the following items and other information:</p> <ol style="list-style-type: none"> <li><input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.</li> <li><input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.</li> <li><input type="checkbox"/> This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).</li> <li><input type="checkbox"/> A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.</li> <li><input type="checkbox"/> copy of the International Application as filed (35 U.S.C. 371 (c)(2)) <ol style="list-style-type: none"> <li><input type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau).</li> <li><input type="checkbox"/> has been transmitted by the International Bureau.</li> <li><input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US).</li> </ol> </li> <li><input checked="" type="checkbox"/> A translation of the International Application into English (35 U.S.C. 371(c)(2))</li> <li><input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) <ol style="list-style-type: none"> <li><input type="checkbox"/> are transmitted herewith (required only if not transmitted by the International Bureau).</li> <li><input type="checkbox"/> have been transmitted by the International Bureau.</li> <li><input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired.</li> <li><input type="checkbox"/> have not been made and will not be made.</li> </ol> </li> <li><input type="checkbox"/> A translation of the amendment to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).</li> <li><input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</li> <li><input type="checkbox"/> A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</li> </ol> <p>Items 11. to 16. below concern document(s) or information included:</p> <ol style="list-style-type: none"> <li><input checked="" type="checkbox"/> An Information Disclosure Statement under 37 C.F.R. 1.97 and 1.98.</li> <li><input checked="" type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 C.F.R. 3.28 and 3.31 is included.</li> <li><input type="checkbox"/> A FIRST preliminary amendment. <input type="checkbox"/> A SECOND OR SUBSEQUENT preliminary amendment.</li> <li><input type="checkbox"/> A substitute specification.</li> <li><input checked="" type="checkbox"/> A change of power of attorney and/or address letter.</li> <li><input checked="" type="checkbox"/> Other items or information:  <u>1</u> Sheets of Drawings  <input checked="" type="checkbox"/> Authorization Pursuant to 37 CFR § 1.136(a)(3) and to Charge Deposit Account </li> </ol>		

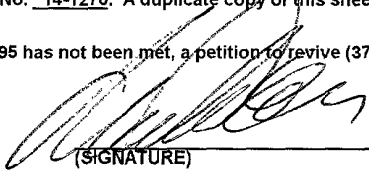
CERTIFICATE OF MAILING

[X] Express Mail Mailing Label No. **EL68694867445**  
Date of Deposit **6-27-01**

I hereby certify that this paper and fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

Jeanne Rusciano  
Typed Name

Jeanne Rusciano  
Signature

U.S. APPLICATION NO. (If known, see 37 C.F.R. 1.5) <b>097/869386</b>		INTERNATIONAL APPLICATION NO. PCT /EP00/10713		ATTORNEY'S DOCKET NUMBER PHN 17,717	
17 [ X ] The following fees are submitted: BASIC NATIONAL FEE (37 C.F.R. 1.492(A)(1)-(5)):				CALCULATIONS (PTO USE ONLY)	
Search Report has been prepared by the EPO or JPO \$860.00 International preliminary-examination fee paid to USPTO (37 C.F.R. 1.482) \$690.00 No international preliminary examination fee paid to USPTO (37 C.F.R. 1.482) but international search fee paid to USPTO (37 C.F.R. 1.445(a)(2)) \$750.00 Neither international preliminary examination fee (37 C.F.R. 1.482) nor international search fee (37 C.F.R. 1.445(a)(2)) paid to USPTO \$970.00 International preliminary examination fee paid to USPTO (37 C.F.R. 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4) \$ 96.00					
ENTER APPROPRIATE BASIC FEE AMOUNT =				\$ 860.00	
Surcharge of \$130.00 for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 C.F.R. 1.492(e)).				\$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	10 - 20 =		X \$ 18.00	\$	
Independent claims	2 - 3 =		X \$ 80.00	\$	
MULTIPLE DEPENDENT CLAIMS (if applicable)			+ \$270.00	\$	
TOTAL OF ABOVE CALCULATIONS =				\$ 860.00	
Reductions by 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed (Note 37 C.F.R. 1.9, 1.27, 1.28)				\$	
SUBTOTAL =				\$ 860.00	
Processing fee of \$130.00 for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 C.F.R. 1.492(f)).				\$	
TOTAL NATIONAL FEE =				\$ 860.00	
Fee for recording the enclosed assignment (37 C.F.R. 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 C.F.R. 3.28, 3.31). \$40.00 per property				\$ 40.00	
TOTAL FEES ENCLOSED =				\$900.00	
				Amount to be refunded	\$
				charged	\$
a. [ ] A check in the amount \$_____ to cover the above fees is enclosed. b. [ X ] Please charge my Deposit Account No. <u>14-1270</u> in the amount of \$ <u>900.00</u> to cover the above fees. A duplicate copy of this sheet is enclosed. c. [ X ] The Commissioner is hereby authorized to charge any additional fee, with the exception of the Base Issue Fee, which may be required, or credit any overpayment to Deposit Account No. <u>14-1270</u> . A duplicate copy of this sheet is enclosed.					
NOTE: Where an appropriate time limit under 37 C.F.R. 1.494 or 1.495 has not been met, a petition to revive (37 C.F.R. 1.137(a) or (b)) must be filed and granted to restore the application to pending status.					
SEND ALL CORRESPONDENCE TO: Corporate Patent Counsel Philips Electronics North America Corporation 580 White Plains Road Tarrytown, NY 10591					
 (SIGNATURE) Michael E. Marion (NAME) 32,266 (REGISTRATION NUMBER)					

1/PR7S

09/869386  
JC18 Rec'd PCT/PTO 27 JUN 2001  
29.10.1999

1

Elimination of noise from a speech signal.

The invention relates to a method for reducing noise in a noisy time-varying input signal, such as a speech signal. The invention further relates to an apparatus for reducing noise in a noisy time-varying input signal.

5 The presence of noise in a time-varying input signal hinders the accuracy and quality of processing the signal. This is particularly the case for processing a speech signal, such as for instance occurs when a speech signal is encoded. The presence of noise is even more destructive if the signal is ultimately not presented to a user, who can relatively well cope with the presence of noise, but if the signal is ultimately processed automatically, as for instance is the case with a speech signal that is recognized automatically. Increasingly automatic speech recognition and coding systems are used. Although the performance of such systems is continuously improving, it is desired that the accuracy be increased further, particularly in adverse environments, such as having a low signal-to-noise ratio (SNR) or a low bandwidth signal. Normally, speech recognition systems compare a representation of an input speech signal against a model  $A_x$  of reference signals, such as hidden Markov models (HMMs) built from representations of a training speech signal. The representations are usually observation vectors with LPC or cepstral components.

10 In practice a mismatch exists between the conditions under which the reference signals (and thus the models) were obtained and the input signal conditions. The reference signals are usually relatively clean (high SNR, high bandwidth), whereas the input signal during actual use is distorted (lower SNR, and/or lower bandwidth). It is, therefore, desired to eliminate at least part of the noise present in the input signal in order to obtain a noise-suppressed signal.

15 A conventional way of estimating a noise-suppressed speech signal ('clean' speech) is to use a spectral subtraction technique. In the discrete-time domain, noise speech  $y$  can be represented as:

$$y(i) = s(i) + n(i), \quad 0 \leq i \leq T-1, \quad (1)$$

20 where  $s$ ,  $n$ ,  $y$  denote clean speech, noise and noisy speech respectively, and where  $T$  denotes the length of the speech and  $i$  is the time index. The conventional spectral subtraction

technique involves determining the spectral components of the noisy-speech and estimating the spectral components of the noise. The spectral components may, for instance, be calculated using a Fast Fourier transform (FFT). The noise spectral components may be estimated once from a part of a signal with predominantly representative noise. Preferably, the noise is

5 estimated 'on-the-fly', for instance each time a 'silent' part is detected in the input signal with no significant amount of speech signal. In the general spectral subtraction technique, the noise-suppressed speech is estimated by subtracting an average noise spectrum from the noisy speech spectrum:

$$\hat{S}(w; m) = \left[ |Y(w; m)|^a - |N(w; m)|^a \right]^{\frac{1}{a}} e^{j\varphi_y(w; m)} \quad (2)$$

10 where  $\hat{S}(w; m)$ ,  $Y(w; m)$ , and  $N(w; m)$  are the magnitude spectrums of the estimated speech  $s$ , noisy speech  $y$  and noise  $n$ ,  $w$  and  $m$  are the frequency and time indices, respectively. The case of  $a = 2$  is referred to as power spectral subtraction. The subtraction is usually called magnitude spectral subtraction if  $a = 1$ .

Due to the subtraction, the estimated spectrum is not guaranteed to be positive in the conventional spectral subtraction techniques. US 5,749,068 describes setting those spectral components to zero for which the subtraction yields a negative outcome:

$$\hat{S}(w) = \max\{Y(w) - \alpha \cdot N(w), 0\} \quad (3)$$

Setting the spectral components to zero (or a low default value) is referred to as 'taking floor' for the negative spectral components. The parameter  $\alpha$ , with a positive value, designates the degree of eliminating noise components. US 5,749,068 describes an advanced way of

20 estimating the spectral components of the noise, but still the conventional spectral subtraction of equation (3) is used.

Taking floor for negative spectral components provides a major limitation of spectral subtraction techniques, introducing residual noise with musical tone artifacts into the

25 estimated speech.

In order to investigate the limitation of the conventional spectrum subtraction techniques, the inventor has carried out an experiment for calculating the ratio of negative spectrum (i.e. the relative number of spectral components which would have a negative value). The negative spectrum ratio  $NSR_{con}$  for the conventional spectral subtraction technique is

30 defined as follows:

$$NSR_{con} = \frac{1}{M} \sum_{k=0}^{M-1} f_{NS} \left( \left| Y(k) \right|^a - \left| \hat{N}(k) \right|^a \right) \tag{4}$$

$$f_{NS}(x) = \begin{cases} 1 & x < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where  $|Y(k)|$  is the corresponding magnitude spectrum of the testing speech  $y$ ,  $|\hat{N}(k)|$  is the noise spectrum estimated from a pause (non-speech segment),  $k$  denotes the  $k$ -th spectrum component and  $M$  represents the total number of spectral components over which the ratio is determined, for instance the number of spectral components in one frame or in the whole testing utterance.

The following table gives the negative spectrum ratio  $NSR_{con}$  for various signal to noise ratios (SNRs) with  $a=2$ . It has been found that the negative spectrum ratio  $NSR_{con}$  even reaches 34.6% at clean conditions. This illustrates that particularly at higher SNR level the conventional spectral subtraction technique introduces some residual noise, limiting the use of the technique.

SNR (dB)	Negative Spectrum Ratio (NSR <sub>con</sub> ) (%)
Clean	34.6
40	22.4
35	18.7
30	14.6
25	10.7
20	7.3
15	4.5
10	2.4
5	1.0
0	0.2

It is an object of the invention to overcome the limitation of the conventional spectral subtraction technique.

To meet the object of the invention, the method for reducing noise in a noisy time-varying input signal  $y$ , such as a speech signal, includes:

receiving the noisy time-varying input signal;

deriving from the signal a plurality of spectral component signals representing respective magnitudes of spectral components of the input signal;

obtaining a correlation coefficient  $\gamma_{sn}$  indicative of a correlation in the spectral domain between a clean speech signal component  $s$  and a noise signal component  $n$  present in the input signal ( $y = s + n$ ); and

estimating magnitudes of respective noise-suppressed spectral components  $\hat{S}(k)$  by solving an equation giving a relationship between the magnitudes of the respective spectral components  $|Y(k)|$  of the noisy input signal  $y$ , the spectral components  $|S(k)|$  of the clean speech signal  $s$ , and the spectral components  $|N(k)|$  of the noise signal  $n$ , where the equation includes the correlation based on the obtained correlation coefficient  $\gamma_{sn}$ . Preferably, the correlation equation is given by:

$$|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn} |S(k)| |N(k)|$$

where  $a$  could be 1 or 2 for magnitude and power spectrum, respectively. Instead of a conventional spectral subtraction this equation is solved which is based on a correlation coefficient  $\gamma_{sn}$  between the clean speech  $s$  and the noise  $n$  in the spectral domain. Solving the equation can be seen as 'correlated spectral subtraction' (CSS).

The correlation coefficient  $\gamma_{sn}$  may be fixed, for instance based on analyzing representative input signals. Preferably, the correlation coefficient  $\gamma_{sn}$  is estimated based on the actual input signal. Advantageously, the estimation is based on minimizing a negative spectrum ratio. Preferably, the expected negative spectrum ratio  $R$  is defined as:

$$R = E\{f_{ns}\} = \frac{1}{M} \sum_{k=0}^{M-1} f_{ns} \left( |Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn} |\hat{S}(k)| |\hat{N}(k)| \right)$$

where advantageously the 'zero-one' function  $f_{ns}$  is given by the differentiable function:

$$f_{ns}(x) = \frac{1}{1 + \exp(-\alpha \cdot x + \beta)}$$

By applying the theory of adaptive learning algorithm, the correlation coefficient is advantageously obtained by following gradient operation.

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta \nabla R$$

The correlation coefficient can be learned along the direction of NSR decrement. Preferably, this is done in an iterative algorithm.

The equation representing the correlated spectral subtraction may be solved directly. Preferably, the equation is solved in an iterative manner, improving the estimate of the clean speech.

5

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments shown in the drawings.

The figure shows a block diagram of a conventional speech processing system wherein the invention can be used.

10

General description of a speech recognition system

The noise reduction according to the invention is particularly useful for processing noisy speech signals, such as coding such a signal or automatically recognizing such a signal. Here a general description of a speech recognition system is given. A person skilled in the art can equally well apply the noise elimination technique in a speech coding system.

Speech recognition systems, such as large vocabulary continuous speech recognition systems, typically use a collection of recognition models to recognize an input pattern. For instance, an acoustic model and a vocabulary may be used to recognize words and a language model may be used to improve the basic recognition result. The figure illustrates a typical structure of a large vocabulary continuous speech recognition system 100. The following definitions are used for describing the system and recognition method:

$A_x$ : a set of trained speech models

$X$ : the original speech which matches the model,  $A_x$

$Y$ : the testing speech

$A_y$ : the matched models for testing environment

$W$ : the word sequence

$S$ : the decoded sequences that can be words, syllables, sub-word units, states or mixture components, or other suitable representations.

The system 100 comprises a spectral analysis subsystem 110 and a unit matching subsystem 120. In the spectral analysis subsystem 110 the speech input signal (SIS) is spectrally and/or temporally analyzed to calculate a representative vector of features (observation vector, OV). Typically, the speech signal is digitized (e.g. sampled at a rate of 6.67 kHz.) and pre-processed, for instance by applying pre-emphasis. Consecutive samples are

grouped (blocked) into frames, corresponding to, for instance, 32 msec. of speech signal. Successive frames partially overlap, for instance, 16 msec. Often the Linear Predictive Coding (LPC) spectral analysis method is used to calculate for each frame a representative vector of features (observation vector). The feature vector may, for instance, have 24, 32 or 63 components. The standard approach to large vocabulary continuous speech recognition is to assume a probabilistic model of speech production, whereby a specified word sequence  $W = w_1 w_2 w_3 \dots w_q$  produces a sequence of acoustic observation vectors  $Y = y_1 y_2 y_3 \dots y_T$ . The recognition error can be statistically minimized by determining the sequence of words  $w_1 w_2 w_3 \dots w_q$  which most probably caused the observed sequence of observation vectors  $y_1 y_2 y_3 \dots y_T$  (over time  $t=1, \dots, T$ ), where the observation vectors are the outcome of the spectral analysis subsystem 110. This results in determining the maximum a posteriori probability:

$$\max P(W|Y, \Lambda_x), \text{ for all possible word sequences } W$$

By applying Bayes' theorem on conditional probabilities,  $P(W|Y, \Lambda_x)$  is given by:

$$P(W | Y, \Lambda_x) = \frac{P(Y | W, \Lambda_x) \cdot P(W)}{P(Y)}$$

Since  $P(Y)$  is independent of  $W$ , the most probable word sequence is given by:

$$\hat{W} = \arg \max_w P(Y, W | \Lambda_x) = \arg \max_w P(Y | W, \Lambda_x) \cdot P(W) \quad (a)$$

In the unit matching subsystem 120, an acoustic model provides the first term of equation (a). The acoustic model is used to estimate the probability  $P(Y|W)$  of a sequence of observation vectors  $Y$  for a given word string  $W$ . For a large vocabulary system, this is usually performed by matching the observation vectors against an inventory of speech recognition units. A speech recognition unit is represented by a sequence of acoustic references. Various forms of speech recognition units may be used. As an example, a whole word or even a group of words may be represented by one speech recognition unit. A word model (WM) provides for each word of a given vocabulary a transcription in a sequence of acoustic references. In most small vocabulary speech recognition systems, a whole word is represented by a speech recognition unit, in which case a direct relationship exists between the word model and the speech recognition unit. In other small vocabulary systems, for instance used for recognizing a relatively large number of words (e.g. several hundreds), or in large vocabulary systems, use can be made of linguistically based sub-word units, such as phones, diphones or syllables, as well as derivative units, such as fenenes and fenones. For such systems, a word model is given by a lexicon 134, describing the sequence of sub-word units relating to a word of the vocabulary, and the sub-word models 132, describing sequences of



acoustic references of the involved speech recognition unit. A word model composer 136 composes the word model based on the sub-word model 132 and the lexicon 134. The (sub-)word models are typically based in Hidden Markov Models (HMMs), which are widely used to stochastically model speech signals. Using such an approach, each recognition unit (word model or sub-word model) is typically characterized by an HMM, whose parameters are estimated from a training set of data. For large vocabulary speech recognition systems usually a limited set of, for instance 40, sub-word units is used, since it would require a lot of training data to adequately train an HMM for larger units. An HMM state corresponds to an acoustic reference. Various techniques are known for modeling a reference, including discrete or continuous probability densities. Each sequence of acoustic references which relate to one specific utterance is also referred as an acoustic transcription of the utterance. It will be appreciated that if other recognition techniques than HMMs are used, details of the acoustic transcription will be different.

A word level matching system 130 of The figure matches the observation vectors against all sequences of speech recognition units and provides the likelihoods of a match between the vector and a sequence. If sub-word units are used, constraints can be placed on the matching by using the lexicon 134 to limit the possible sequence of sub-word units to sequences in the lexicon 134. This reduces the outcome to possible sequences of words.

Furthermore a sentence level matching system 140 may be used which, based on a language model (LM), places further constraints on the matching so that the paths investigated are those corresponding to word sequences which are proper sequences as specified by the language model. As such the language model provides the second term  $P(W)$  of equation (a). Combining the results of the acoustic model with those of the language model, results in an outcome of the unit matching subsystem 120 which is a recognized sentence (RS) 152. The language model used in pattern recognition may include syntactical and/or semantical constraints 142 of the language and the recognition task. A language model based on syntactical constraints is usually referred to as a grammar 144. The grammar 144 used by the language model provides the probability of a word sequence  $W = w_1 w_2 w_3 \dots w_q$ , which in principle is given by:

$$P(W) = P(w_1)P(w_2|w_1).P(w_3|w_1w_2)...P(w_q|w_1w_2w_3...w_q).$$

Since in practice it is infeasible to reliably estimate the conditional word probabilities for all words and all sequence lengths in a given language, N-gram word models are widely used. In an N-gram model, the term  $P(w_j|w_1w_2w_3...w_{j-1})$  is approximated by  $P(w_j|w_{j-N+1}...w_{j-1})$ . In

practice, bigrams or trigrams are used. In a trigram, the term  $P(w_j | w_1 w_2 w_3 \dots w_{j-1})$  is approximated by  $P(w_j | w_{j-2} w_{j-1})$ .

The speech processing system according to the invention may be implemented using conventional hardware. For instance, a speech recognition system may be implemented on a computer, such as a PC, where the speech input is received via a microphone and digitized by a conventional audio interface card. All additional processing takes place in the form of software procedures executed by the CPU. In particular, the speech may be received via a telephone connection, e.g. using a conventional modem in the computer. The speech processing may also be performed using dedicated hardware, e.g. built around a DSP.

The noise elimination according to the invention may be performed in a pre-processing step before the spectral analysis subsystem 100. Preferably, the noise elimination is integrated in the spectral analysis subsystem 100, for instance to avoid that several conversions from the time domain to the spectral domain and vice versa are required. All hardware and processing capabilities for performing the invention are normally present in a speech recognition or speech coding system. The noise elimination technique according to the invention is normally executed on a processor, such as a DSP or microprocessor of a personal computer, under control of a suitable program. Programming the elementary functions of the noise elimination technique, such as performing a conversion from the time domain to the spectral domain, falls well within the range of a skilled person.

Detailed description of the invention

Details are given for speech signals. Other signals can be processed in a corresponding way. As described above, in the discrete-time domain noise speech  $y$  can be represented as:

$$y(i) = s(i) + n(i), \quad 0 \leq i \leq T-1, \quad (1)$$

where  $s$ ,  $n$ ,  $y$  denote clean speech, noise and noisy speech respectively, and where  $T$  denotes the length of the speech and  $i$  is the time index. Using conventional techniques, such as a Fast Fourier transform, the speech signal  $y$  can be transformed into a set of spectral components  $Y(k)$ . It will be appreciated that if already a suitable conversion to the time domain had taken place, it is sufficient to retrieve the spectral components resulting from such a conversion.

Let  $|S(k)|$ ,  $|N(k)|$ , and  $|Y(k)|$  be the corresponding magnitude of the spectrums of the time-domain signals  $s$ ,  $n$ , and  $y$ , respectively. Using the conventional spectral subtraction techniques, individual spectral components are forced to be positive. It does not

allow the situation wherein an individual spectral component  $Y(k)$  of the noisy speech  $y$  is less than the corresponding spectral component  $N(k)$  of the noise signal  $n$ .

The following correlation is assumed to exist between the speech signal and the noise signal:

$$|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn} |S(k)| |N(k)| \quad (6)$$

where  $\gamma_{sn}$  denotes the correlation coefficient of speech and noise in the spectral domain and  $a$  could be 1 or 2 for magnitude and power spectrum, respectively. Using this correlation as the basis for estimating the clean speech spectrum (and as such using a correlated spectral subtraction) makes it possible to have the situation wherein  $|Y(k)|^a < |N(k)|^a$  if  $\gamma_{sn} < 0$ .

Let  $|\hat{s}(k)|$  and  $|\hat{n}(k)|$  be the estimates of the magnitude spectrums of the clean speech signal  $s$  and the noise signal  $n$ , respectively. Preferably,  $|\hat{n}(k)|$  is estimated from pause (non-speech segment). Based on equation (6),  $|\hat{s}(k)|$  can be calculated by solving the equation in one step or by using an iterative algorithm. The one-step solution are give in the following equations (7) and (8) for the cases wherein  $a=1$  or  $a=2$ , respectively:

$$|\hat{s}(k)| = \frac{|Y(k)| - |N(k)|}{(1 + \gamma_{sn} |N(k)|)}, \text{ if } a=1 \quad (7)$$

$$|\hat{s}(k)| = \frac{-\gamma_{sn} |N(k)| \pm \sqrt{\gamma_{sn}^2 |N(k)|^2 + 4(|Y(k)|^2 - |N(k)|^2)}}{2}, \text{ if } a=2 \quad (8)$$

Equation (8) has two possible solutions. The positive solution which is greater than  $(|Y(k)|^2 - |N(k)|^2)$  or close to  $(|Y(k)|^2 - |N(k)|^2)$  will be chosen since the direction of NSR decrement is preferred.

A preferred iterative algorithm for estimating  $|\hat{s}(k)|$  with specified correlation coefficient,  $\gamma_{sn}$ , is as follows:

LOOP k ( 0 : N-1 )

$$\text{Initialization: } \left| \hat{S}^{(0)}(k) \right|^a = \left| Y(k) \right|^a - \left| \hat{N}(k) \right|^a \quad (9)$$

LOOP  $\ell$

$$5 \quad \left| S'(k) \right|^a = \left| Y(k) \right|^a - \left| \hat{N}(k) \right|^a - 2\gamma_{sn} \left| \hat{S}^{(\ell)}(k) \right| \left| \hat{N}(k) \right| \quad (10)$$

$$\left| \hat{S}^{(\ell+1)}(k) \right|^a = \frac{\left| S'(k) \right|^a + \left| \hat{S}^{(\ell)}(k) \right|^a}{2} \quad (11)$$

$$\text{IF } \frac{\left| \left| \hat{S}^{(\ell+1)}(k) \right|^a - \left| \hat{S}^{(\ell)}(k) \right|^a \right|}{\left| \hat{S}^{(\ell)}(k) \right|^a} < \text{Threshold THEN STOP}$$

ELSE  $\ell = \ell + 1$

END LOOP  $\ell$

10 END LOOP k

The outer loop k deals with all individual spectral components. The inner loop is performed until the iteration has converged (no significant change occurs anymore in the estimated speech).

15 The above described algorithm can be used for a fixed correlation coefficient  $\gamma_{sn}$ . In a further embodiment according to the invention, the correlation coefficient  $\gamma_{sn}$  is estimated based on the actual input signal y. To this end, the function of negative spectrum ratio (NSR) for the correlated spectral subtraction algorithm according to the invention is defined as follows:

$$20 \quad NSR = \frac{1}{M} \sum_{k=0}^{M-1} f_{NS} \left( \left| Y(k) \right|^a - \left| \hat{N}(k) \right|^a - \gamma_{sn} \left| \hat{S}(k) \right| \left| \hat{N}(k) \right| \right) \quad (12)$$

The  $f_{NS}$  function shown in equation (5) is a zero-one function. In order to derive the relation between the correlation coefficient  $\gamma_{sn}$  and NSR, a smoothed zero-one, sigmoid function family is preferably used. For example, the following function  $f_{ns}$  is advantageously used for further derivation due to its differentiability.

$$25 \quad f_{ns}(x) = \frac{1}{1 + \exp(-\alpha \cdot x + \beta)} \quad (13)$$

Exemplary values for  $\alpha$  and  $\beta$  are 1.0 and 0.0, respectively.

Then, the expected negative spectrum ratio  $R$  is defined as follows:

$$R = E\{f_{ns}\} = \frac{1}{M} \sum_{k=0}^{M-1} f_{ns} \left( |Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn} |\hat{S}(k)| |\hat{N}(k)| \right) \quad (14)$$

By applying the theory of adaptive learning algorithm, the correlation coefficient is preferably

5 obtained by the following gradient operation:

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta \nabla R \quad (15)$$

The correlation coefficient can be learned along the direction of decrease in NSR. This implies to reduce the residual noise in the estimated spectrum using the proposed correlated spectral subtraction (CSS) algorithm.

10 The algorithm of estimating  $|\hat{S}(k)|$  with a minimum NSR based correlation coefficient  $\gamma_{sn}$  is as follows:

Initialization:  $m = 0$

$\gamma_{sn}^{(m)}$  = non-zero, initial guess.

15 LOOP m

LOOP k (0 : N-1)

*Block I*

$\ell = 0$

$$|\hat{S}^{(\ell)}(k)|^a = |Y(k)|^a - |\hat{N}(k)|^a$$

LOOP  $\ell$

$$|S'(k)|^a = |Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn}^{(m)} |\hat{S}^{(\ell)}(k)| |\hat{N}(k)|$$

$$|\hat{S}^{(\ell+1)}(k)|^a = \frac{|S'(k)|^a + |\hat{S}^{(\ell)}(k)|^a}{2}$$

If  $\frac{|\hat{S}^{(\ell+1)}(k)|^a - |\hat{S}^{(\ell)}(k)|^a}{|\hat{S}^{(\ell)}(k)|^a} < \text{Threshold 1}$  THEN STOP

ELSE  $\ell = \ell + 1$

END LOOP  $\ell$

25 END LOOP k

$$R = E\{f_{ns}\} = \frac{1}{M} \sum_{k=0}^{M-1} f_{ns} \left( |Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn} |\hat{S}(k)| |\hat{N}(k)| \right)$$

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta \nabla R$$

$$\text{If } \left| \frac{\gamma_{sn}^{(m+1)} - \gamma_{sn}^{(m)}}{\gamma_{sn}^{(m)}} \right| < \text{Threshold 2 THEN STOP}$$

END LOOP m

5

The block indicated as block 1 is the same as used for the iterative algorithm assuming a fixed correlation coefficient  $\gamma_{sn}$ . Instead of using the iterative solution in block one, also the one-step solution of equations (7) or (8) may be used.

It will be appreciated that after the noise has been eliminated as described above, the resulting estimated spectral components of the noise-eliminated signal may be converted back to the time-domain. Where possible the spectral components may be used directly for the subsequent further processing, like coding or automatically recognizing the signal.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

## CLAIMS:

1. A method for reducing noise in a noisy time-varying input signal  $y$ , such as a speech signal; the method including:
  - receiving the noisy time-varying input signal  $y$ ;
  - deriving from the input signal  $y$  a plurality of spectral component signals
  - representing respective magnitudes  $|Y(k)|$  of spectral components of the input signal  $y$ ;
  - obtaining a correlation coefficient  $\gamma_{sn}$  indicative of a correlation in the spectral domain between a clean speech signal component  $s$  and a noise signal component  $n$  present in the input signal  $y$  ( $y = s + n$ ); and
  - estimating magnitudes of respective noise-suppressed spectral components  $\hat{S}(k)$  by solving a correlation equation giving a relationship between the magnitudes of the respective spectral components  $|Y(k)|$  of the noisy input signal  $y$ , the spectral components  $|S(k)|$  of the clean speech signal  $s$ , and the spectral components  $|N(k)|$  of the noise signal  $n$ , where the equation includes the correlation based on the obtained correlation coefficient  $\gamma_{sn}$ .
2. The method as claimed in claim 1, wherein the correlation coefficient  $\gamma_{sn}$  is predetermined.
3. The method as claimed in claim 1, wherein the step of obtaining the correlation coefficient  $\gamma_{sn}$  includes estimating the correlation coefficient  $\gamma_{sn}$ .
4. The method as claimed in claim 3, wherein the step of estimating the correlation coefficient  $\gamma_{sn}$  includes determining a minimum negative spectrum ratio.
5. The method as claimed in claim 4, wherein the negative spectrum ratio NSR represents a proportion of spectral components  $\hat{S}(k)$  which would be negative based on the solution of the correlation equation.
6. The method as claimed in claim 5, wherein the method includes:
  - initializing the correlation coefficient  $\gamma_{sn}$  with a non-zero value; and

iteratively:

performing the step of solving the correlation equation to obtain  $|\hat{S}(k)|$ ;

and

estimating a new correlation coefficient based on a gradient decent of

5 the negative spectrum ratio NSR for  $\hat{S}(k)$ .

7. The method as claimed in claim 1, wherein the step of solving the correlation equation includes iteratively estimating the noise-suppressed spectrum  $\hat{S}(k)$ .

10 8. The method as claimed in claim 7, wherein method includes calculating an initial estimate of a magnitude of the noise-suppressed spectrum  $\hat{S}^{(0)}(k)$  by subtracting a magnitude of an estimate of the respective spectral components  $\hat{N}(k)$  of the noise signal  $n$  from a magnitude of the respective spectral components  $Y(k)$  of the noisy input signal  $y$ .

15 9. The method as claimed in claim 7, wherein the step of performing the iterative spectrum estimation includes in each iteration:

estimating a magnitude of an auxiliary noise-suppressed spectrum based on the correlation equation where a term with the correlation coefficient  $\gamma_{sn}$  is based on a current estimate of a magnitude of the noise-suppressed spectrum  $\hat{S}^{(l)}(k)$ ; and

20 estimating a new magnitude of the noise-suppressed spectrum  $\hat{S}^{(l+1)}(k)$  based the estimated magnitude of the auxiliary noise-suppressed spectrum and on the current estimate of a magnitude of the noise-suppressed spectrum  $\hat{S}^{(l)}(k)$ .

25 10. An apparatus for reducing noise in a noisy time-varying input signal  $y$ , such as a speech signal; the apparatus including:

an input for receiving the noisy time-varying input signal  $y$ ;

means for deriving from the input signal  $y$  a plurality of spectral component signals representing respective magnitudes  $|Y(k)|$  of spectral components of the input signal  $y$ ;

30 means for obtaining a correlation coefficient  $\gamma_{sn}$  indicative of a correlation in the spectral domain between a clean speech signal component  $s$  and a noise signal component  $n$  present in the input signal  $y$  ( $y = s + n$ ); and



means for estimating magnitudes of respective noise-suppressed spectral components  $\hat{S}(k)$  by solving a correlation equation giving a relationship between the magnitudes of the respective spectral components  $|Y(k)|$  of the noisy input signal  $y$ , the spectral components  $|S(k)|$  of the clean speech signal  $s$ , and the spectral components  $|N(k)|$  of the noise signal  $n$ , where the equation includes the correlation based on the obtained correlation coefficient  $\gamma_{sn}$ .

102250 9362000

## ABSTRACT:

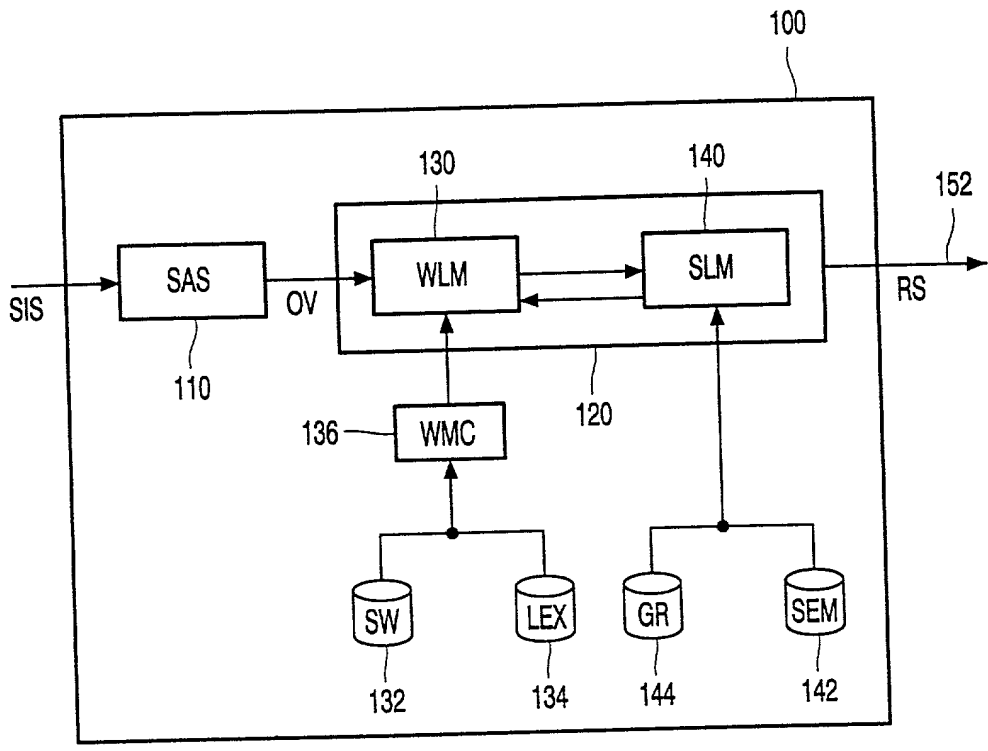
A method for reducing noise in a noisy time-varying speech input signal  $y$  includes receiving the input signal  $y$  and deriving a plurality of spectral component signals representing respective magnitudes  $|Y(k)|$  of spectral components of the input signal  $y$ .

A correlation coefficient  $\gamma_{sn}$  is obtained which indicates a correlation in the spectral domain between a clean speech signal component  $s$  and a noise signal component  $n$  present in the input signal  $y$  ( $y = s + n$ ).

Magnitudes of respective noise-suppressed spectral components  $\hat{S}(k)$  are estimated by solving a correlation equation which gives a relationship between the magnitudes of the respective spectral components  $|Y(k)|$  of the noisy input signal  $y$ , the spectral components  $|S(k)|$  of the clean speech signal  $s$ , and the spectral components  $|N(k)|$  of the noise signal  $n$ , where the equation includes the correlation based on the obtained correlation coefficient  $\gamma_{sn}$ . Preferably, the correlation equation is given by

$$|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn} |S(k)| |N(k)|.$$

Fig.



# DECLARATION and POWER OF ATTORNEY

ATTORNEY'S DOCKET NO.:  
**PHN 17.717**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**"Elimination of noise from a speech signal"**

the specification of which (check one)

☒ is attached hereto.

☐ was filed on \_\_\_\_\_ as Application Serial No. \_\_\_\_\_ and was amended on \_\_\_\_\_ (if

applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by the amendment(s) referred to above.

I acknowledge the duty to disclose information which is material to patentability of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

## PRIOR FOREIGN APPLICATION(S)

COUNTRY	APP. NUMBER	DATE OF FILING (DATE, MONTH, YEAR)	PRIORITY CLAIMED UNDER 35 U.S.C. 119
Europe	99203565.9	29 October 1999	YES

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35 United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

## PRIOR UNITED STATES APPLICATION(S)

APPLICATION SERIAL NUMBER	FILING DATE	STATUS (PATENTED, PENDING, ABANDONED)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

**POWER OF ATTORNEY:** As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

3 Jack E. Haken, Reg. No. 26,902

Michael E. Marion, Reg. No. 32,266

Edward M. Blocker, Reg. No. 30,245

SEND CORRESPONDENCE TO: Corporate Patent Counsel; U.S. Philips Corporation; 580 white Plains Road; Tarrytown, NY 10591	DIRECT TELEPHONE CALLS TO: (name and telephone No.) (914) 332-0222
--	--

Dated:		Inventor's Signature: <i>Chao-shih Huang</i>	
Full Name of inventor	Last Name <b>HUANG</b>	First Name <b>Chao-Shih</b>	Middle Name <b>Joseph</b>
Residence & Citizenship	City <b>Tauyan</b>	State of Foreign Country <b>Taiwan TWX</b>	Country of Citizenship <b>Taiwan</b>
Post Office Address	Street <b>No. 10, Lane 1, Alley 541, Wen-Hua St.</b>	City <b>326 Tauyan</b>	State of Country <b>Taiwan</b>
			Zip Code

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF  
ATTORNEY

(includes Reference to PCT International Applications)

ATTORNEY'S DOCKET  
NUMBER

PHN 17.717 US

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **"Elimination of noise from a speech signal"**  
the specification of which (check only one item below):

☒ is attached hereto.

☐ was filed as United States application

Serial No \_\_\_\_\_

on \_\_\_\_\_

and was amended

on \_\_\_\_\_

☒ was filed as PCT international application

Number PCT/EP00/10713

on 27 October 2000

and was amended under PCT Article 19

on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

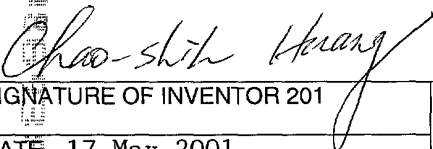
I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, § 1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:

COUNTRY	APPLICATION NUMBER	DATE OF FILING DAY, MONTH, YEAR	PRIORITY CLAIMED UNDER 35 USC 119
Europe	99203565.9	29 October 1999	YES

U.S. DEPARTMENT OF COMMERCE -Patent and Trademarks Office  
(July 1994)

Combined Declaration For Patent Application and Power of Attorney (Continued) (includes Reference to PCT International Applications)				Attorneys Docket Number <b>PHN 17.717 US</b>
POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (List name and registration number)				
Jack E. Haken, Reg. No. 26,902 Michael E. Marion, Reg. 32,266 Edward M. Blocker, Reg. No. 30,245				Direct Telephone Calls to: (name and telephone number) (914)332-0222
201	FULL NAME OF INVENTOR	FAMILY NAME <b>HUANG</b>	FIRST GIVEN NAME <b>Chao-Shih</b>	SECONDE GIVEN NAME <b>Joseph</b>
	RESIDENCE & CITIZENSHIP	CITY <b>Tauyan</b>	STATE OR FOREIGN COUNTRY <b>Taiwan</b>	COUNTRY OF CITIZENSHIP <b>Taiwan</b>
	POST OFFICE ADDRESS	POST OFFICE ADDRESS <b>No. 10, Lane 1, Alley 541, Wen-Hua St.</b>	CITY <b>326 Tauyan</b>	STATE & ZIP CODE/COUNTRY <b>Taiwan</b>
<p>I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true: and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.</p> <div style="margin-top: 20px;">  </div>				
SIGNATURE OF INVENTOR 201				
DATE 17 May 2001				

U.S. DEPARTMENT OF COMMERCE- Patent and Trademarks Office  
(July 1994)